# Prediction Of Raman Spectra and Adulteration Concentration of Wheat Flour Based on Neural Network Modelling

Zongze Li [1], Jingru Xiao [2, *]

[1] Department of College of Software Engineering, Sichuan University, Chengdu, China

[2] School of Chengdu foreign languages school, Chengdu, China

* Corresponding Author Email: lizongze@stu.scu.edu.cn

**Abstract.** In order to quickly detect the adulteration of flour containing wheat flour quality testing, the Raman spectrum of wheat flour as the object of study, based on a number of data-based neural network to identify and determine the concentration of wheat adulteration. Firstly, the neural network system in this study borrowed the Raman wave number and Raman intensity curve from "Non-contact Detection of Benzoyl Peroxide in Flour Based on Raman Spectroscopy" as the training and testing set. Next, we created the neural network, set the training parameters, trained the network and simulated the test. Finally, the supervised learning process resulted in error analysis and comparison of results, and the Raman spectra of wheat flour and adulteration concentration were plotted visually.

**Keywords:** Neural network; raman spectra; deep learning.

## 1. Introduction

Nowadays, wheat flour is going all over the world as the logistics tends to mature. Because wheat flour is one of the most important foods in the whole society, so like for increasing the purity of wheat flour and product quality issues become the focus of great concern. Some unscrupulous merchants have been mixing gypsum powder or fine gravel or metal to increase the weight of wheat flour, which greatly affects the quality of wheat flour and the social atmosphere. However, the traditional experimental testing methods have problems such as large sample size, long testing period and high cost, which are not applicable to the rapid testing of wheat flour on the sales site, making the purity problem of people's concern cannot be solved. In recent years, Raman spectroscopy has been developed as a high-sensitivity rapid analysis method, which is widely used in the field of food adulteration. [1] Its superiority is non-destructive detection, low cost, no pollution and online analysis, which can fill the gap of rapid detection of wheat flour adulteration concentration. This neural network is based on the characteristic peaks of Raman spectra and the corresponding adulteration content of wheat flour in different cases, which was created by the author to achieve rapid identification of adulteration, to better serve consumers, to screen out products with higher purity and healthier, and thus to improve the quality of life. [2][3]

Traditional Raman spectroscopy is a scattering spectrum, which is generated based on the interaction of light and materials. [4] Definition of Raman scattering: is that molecules scattered the high-intensity incident light of a laser source, the scattered light has the same color as the incident laser, also this scattering is called that Rayleigh scattering. Nevertheless, there's a small fraction (about 1/10^9) of the scattered light has a different wavelength from the incident light, and the chemical structure of the test sample (the so-called scattering material) determined wavelength change. This part of the scattered light is called Raman scattering. [5]

In this project, we used the deep learning method to predict the harmful substances in wheat. Deep learning is learning the inherent law of sample data and presentation layers, the information obtained in the process of learning such as text, the interpretation of the data in graphics and sound has a lot of help. [6] Its ultimate goal is for machines to be able to analyze and learn like humans, able to recognize data. [7] Deep learning is a complex machine learning algorithm that has achieved far better results in image recognition and speech than previous techniques. [8][9] Deep learning has achieved many successes in many fields, such as machine learning, machine translation, number mining, natural language processing, and multimedia personalization.[10] This technology provides sufficient conditions for machines to mimic human visual, auditory, and mental activities, so that artificial intelligence achieves the goal of recognizing more complex models and witnesses a step towards maturity in this field of technology.

## 2. Method

### 2.1. Datasets

#### 2.1.1 Data collection

The data is collected from the Raman wave number and Raman intensity curve of 10 mixed concentration samples in the "non-contact detection of benzoyl peroxide in flour based on Raman spectroscopy". [11] The Raman spectrometer is scanned with a scanning range of 400-2200cm-1, and the spectral curve of each sample contains 200 waves. Count the points, and the interval is 10 waves. At the same time, the range of pesticide content is 0.05~50%.

#### 2.1.2 Data filtering

We use getdata graphics digitisation software to take 200 data points for each spectrum, a total of 2,000 data samples. The sample data we refer to is extracted from the Raman spectrum line diagram in the non-contact detection of benzoyl peroxide in flour based on Raman spectroscopy. The X axis is the Raman wave number, and the Y axis is the Raman intensity. On each folding line, we use getdata (a digital software) to describe an average of 200 data points about X and Y. If the data points taken by each line are different, it turns out that these data cannot be trained correctly. Because after the data is transposed into a matrix, it is no longer a square matrix. In addition, in the process of tracing points, it is very important to take points evenly for each line. Otherwise, take more points from 200cm^-1 to 1200cm^-1, and take fewer points from 1201cm^-1 to 2200cm^-1, and the training model is more "deformed". It is very accurate for some data predictions close to the first part, but the deviation for data predictions closer to the second part is large. Obviously, to ensure the accuracy and generalisation of the model, we must average the data points. [12]

#### 2.1.3 Data analysis

After taking 10*200 pieces of data, we first preprocess the original data and convert the data group into a two-dimensional matrix for model training. Then divide the data into two groups for the first simulation test, 80% for training and 20% for testing. We don't have out of bag data (OOB). Finally, the test results are drawn into a line chart, and it is found that there are only two test nodes, and the deviation between the test results and the predicted results is large, and the correlation coefficient is only about 0.6~0.7.)

Therefore, we carry out data enhancement, expand the sample data to 60 and 120, and still divide it into two sets of tests. In the end, the results we tested were very accurate.

### 2.2. Assuming of results

#### 2.2.1 Evaluation Criteria

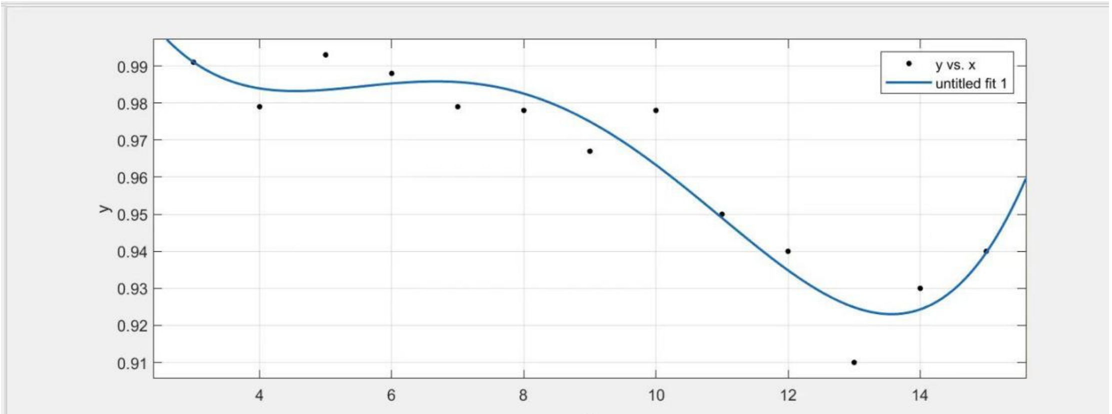


**Fig. 1** Accuracy

Figure 1 shows the accuracy of the prediction model. Accuracy Specify the negative units of the computer's operation to ten when performing data operations. In this study, the results of predicting the doped concentration in the wheat Raman spectrum are calculated to a certain unit. In fact, we use the control variable method to filter for accuracy in order to select the most accurate number of results. We set the accuracy to 1e-3, 1e-5, 1e-8 respectively. Finally, in order to compare the accuracy of the prediction results of different accuracy, we make a line chart of accuracy.

Method of programming:

First clear the lines of code. In the second part, the training set is set. In this experiment, we set the training set parameters to 90, and the test parameters to 50. The third part creates BP neural network creation, and Training network, Simulation test. Section 4 conducts the performance evaluation and finds the relative error, Decision coefficient and Result contrast. The last step is done by drawing using the tools of matlab. [4]

Number of hidden nodes as we all know, the middle layer of BP neural network is a hidden layer, which is used to process information transformation. According to the needs of different accuracy, the number of hidden layer nodes can be changed or designed as a single hidden layer or multi-hidden layer structure. [ 3] The number of nodes in the hidden layer affects the prediction ability and accuracy of the network. To this end, we still use the control variable method to test different variables. After testing the number of 3 to 20 hidden layer nodes, we determined that when the number of hidden nodes is 5, the final prediction result is the best.

### 2.3. Algorithm

BP neural networks do not need to describe the mathematical equations of the mapping relationship in advance, by learning and storing a large number of mapping relationships between input and output modes, and then through back propagation to continuously adjust the weights and thresholds of the network, and ultimately make the neural network to minimize the sum of squared error. [2][4]

(1) Initialize the weights and thresholds with appropriate values.

(2) Obtain the "input" from the training data, train the format {input, correct output}, pass the input to the neural network model output, calculate the difference between the correct output di and the model output.

(3) Calculate the update of weights according to the incremental rule

(4) Adjust the weights

(5) Repeat the steps 2-4

### 2.4. Metric

#### 2.4.1 Root Mean Square Error

Root means square error is the square root of the average of all squared errors.

Root Mean Square Error = sqrt (1/n * ∑(y_pred_i - y_true_i)^2)

This metric allows you to check if the model structure is correct.

#### 2.4.2 Correlation coefficient R^2

The correlation coefficient, usually between 0 and 1, determines the strength and direction of the linear relationship between the predicted and true values of the two random variables.

## 3. Result

### 3.1. Different intensity

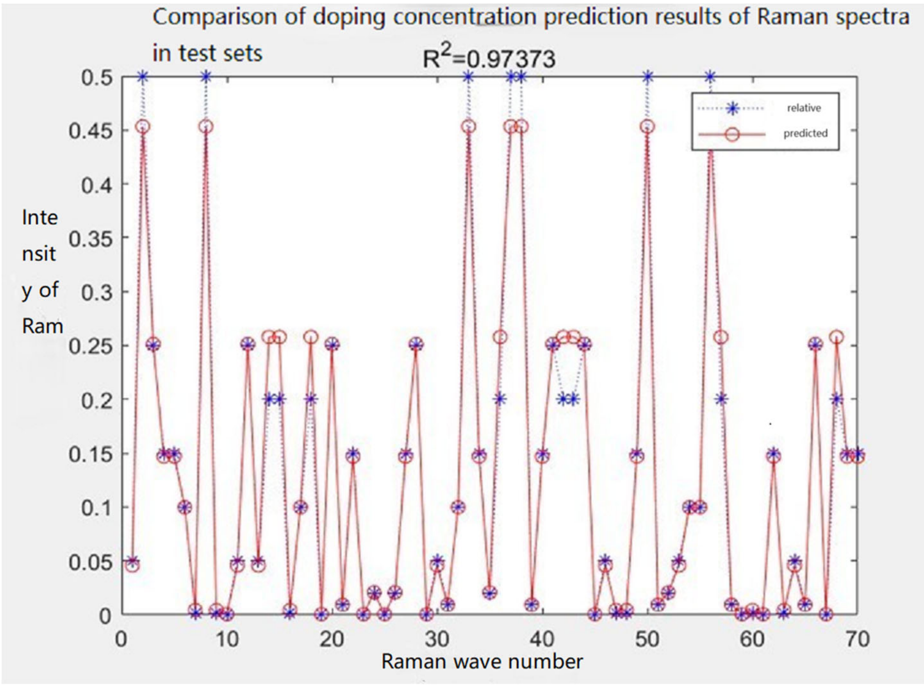


**Fig. 2** Intensity of Ram.

Figure 2 shows the relationship between the predicted Raman strength and the true value at different Raman wave numbers.In this figure, the horizontal axis represents the Raman wave number in cm, the vertical axis represents the Raman intensity, the R-square above represents the coincidence between the real value and the predicted value, the blue line in the chart is the real value, the asterisk above represents the data point, the red line represents the expected value calculated by the program, and the circle above is the predicted value. We find that the Raman wave number is 2cm, the Raman intensity reaches the first peak at 0.25, and higher peaks appear at 17, 22 and 24cm, and the Raman intensity reaches as high as 0.5 at 26 and 28cm, but the predicted value deviates from the real value, while the other bands are generally consistent.

### 3.2. Comparison of hypotheses

For the collected 20 groups of samples with different doping content, Raman spectrometer was used to scan them, and the scanning range was 400-2200cm-1. The spectral curve of each sample contained a total of 180 wave number points, separated by 10 wave numbers. Meanwhile, the pesticide content ranged from 0.05 to 50%.

This experiment through the application of BP neural network to create, training and simulation test, and then through the performance evaluation, and then drawing.

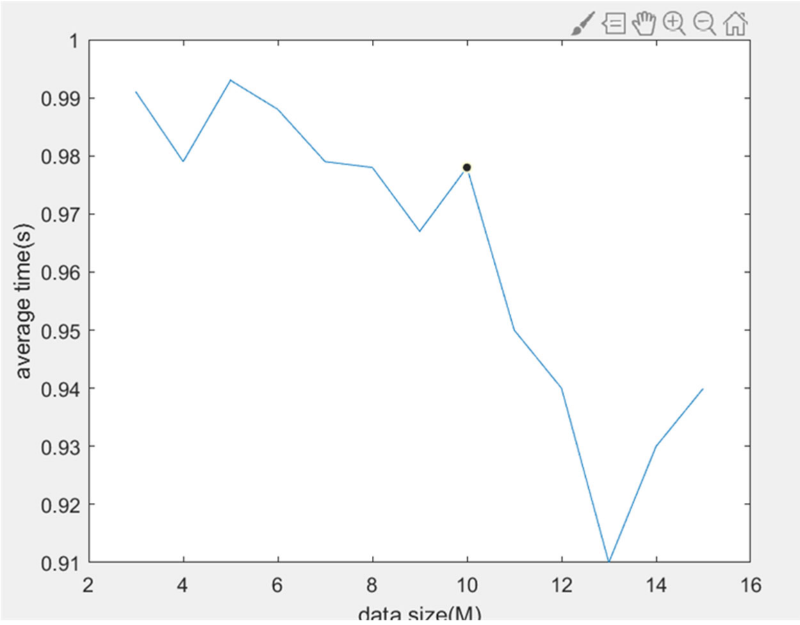


**Fig. 3** Average time.

Figure 3 represents the relationship between the number of hidden layer nodes and R squared. When the number of hidden layer nodes is 5 maximum correlation coefficients indexing as high as 0.993, when the number of hidden layer nodes is 13 minimum correlation coefficient index is 0.910.

## 4. Discussion

We validated a neural network is based model to predict Raman spectra and adulteration concentration of wheat flour. Combining the existing experience and methods of data extraction, data expansion, and data screening, the neural network displays output values corresponding to adulteration when Raman spectral data of adulteration is input. Relying on this neural network, people all over the world can more easily detect whether wheat flour is adulterated or not. By collecting Raman spectra from wheat flour and classifying and debugging the training set with the test set, an efficient neural network learning model can be constructed, which greatly improves the efficiency and accuracy of the detection, and at the same time can guarantee the quality of wheat flour. In addition, it is recommended to randomly delete part of the data from the training and test sets after expanding the data set to avoid excessive accuracy caused by the high degree of overlap between the training and test sets, which makes the experimental results contingent and random. At the same time, the ratio of the data in the training set to the test set can also be appropriately reduced to make the simulation test more accurate; Nevertheless, due to the different types of wheat flour around the world, and the wheat flour adulteration standard judgment varies from country to country, this neural network system still has limitations, one-sidedness, and objectivity, so a large number of different varieties of wheat flour are still needed for the deep learning of the neural system. In addition, there are relatively few portable devices for wheat flour Raman spectroscopy extraction, which makes it difficult to obtain data quickly. If the sampling problem can be solved, the neural network model can be more fully utilized to manage wheat flour adulteration and better protect food safety.

## 5. Conclusion

The results of three groups of experiments show that the coincidence rate of the adulteration concentration of wheat flour predicted by Raman spectroscopy and the adulteration concentration of wheat flour is 99 percent. Illustrates the use of Raman spectrum measurement is effective and accurate.

Experiments of adulterated in wheat flour was smooth import of nervous system, the input adulterated Raman spectra data, the neural network will show the corresponding output values of adulteration. Based on this neural network, people around the world can more easily detect whether wheat flour is adulterated. By collecting the Raman spectra of wheat flour and classifying and debugging the training set and test set, an efficient neural network learning model can be constructed, which greatly improves the efficiency and accuracy of detection and ensures the quality of wheat flour without damage.

However, insufficient number of samples or the insufficient variety of sample sources affect the representativeness and credibility of the research results. Based on the world's wheat starch adulterated test, we are short of samples of different wheat varieties.

In addition, there is a lack of portable detection instruments in daily life. If the detection system of Raman spectroscopy can be improved and the neural network system can be supported, the adulteration detection of wheat flour can be more successful, and the quality of life can be improved.

## Authors Contribution

All the authors contributed equally, and their names were listed in alphabetical order.

## References

[1] Zhao Juan, & Peng Yankun. (2015). Non-contact detection of benzoyl peroxide in flour based on Raman spectroscopy. China Agricultural Science and Technology Herald, (5), 33-39.

[2] Wee, & Zhang, Hua Yue. (2014). Analysis of gold price prediction based on BP neural network. Tianjin Science and Technology, 41(1), 68-73.

[3] Xu Tianfeng, Peng Yankun, Li Yongyu, Zhai Chen, Zheng Xiaochun, & Qiao Lu. (2016). Raman point scanning method for rapid detection of spinach quality and safety parameters. Spectroscopy and Spectral Analysis, 36(6), 1765-1770.

[4] Yanyan Huang, & Jian Yao. (2023). Prediction of rice moisture content based on PSO-GA-BP neural network. Grain and Feed Industry, (3), 56-60.

[5] TONG Quan, SHI Zhixin, & MAO Zhiwei. (2023). Spatter analysis of rotating arc images based on multi-thresholding and neural network. Journal of Welding, 43(12), 41-46.

[6] Peng, Yankun, Tian, Wenjian, Guo, Qinghui, Zhao, Xinlong, & Qiao, Xin. (2021). Non-destructive quantitative analysis of acetamiprid pesticide residues in apples based on enhanced Raman spectroscopy. Transactions of the Chinese Society of Agricultural Engineering, 37(14).

[7] Yuan Feiniu, Zhang Lin, Shi Jinting, Xia Xue, & Li Gang. (2019). A review on the theory and application of self-coding neural networks. Journal of Computing, 42(1), 203-230.

[8] Yang, S. E., & Wang, L. P.. (2007). Early warning of financial crisis of listed companies based on BP neural network and panel data (Doctoral dissertation).

[9] Gan, Jingyi, Huang, Minghe, & Yuan, Jing. (2011). Analysis of China's financial risk early warning model based on BP artificial neural network. Business Times, (27), 65-66.

[10] Yang, J., & Zeng, W.W.. (2022). A prediction model of surface settlement in mining area based on GA-BP neural network. Mining and Metallurgical Engineering, 42(2), 42-45.

[11] Li, W.H., Wen, L., & Chen, Y.B.. (2017). Application of improved GA-BP neural network model in property crime prediction. Journal of Wuhan University (Information Science Edition), 42(8), 1110-1116.

[12] Chen, Yuan-Ling, Wang, Xiao, Sun, Ying-Jie, & Zhang, Yang. (2022). Sugarcane harvest quality prediction based on GA-BP neural network. Agricultural Mechanization Research, 44(2), 187-191.